GMM estimation of lattice models using panel data: application

Théophile Azomahou*

April 2001

BETA-Theme, Université Louis Pasteur, Strasbourg
Department of Economics
61, avenue de la Forêt Noire
F-67085 Strasbourg Cedex; France

Abstract

We propose an empirical application of lattice models to actual household-level data based on the generalized method of moments. We take advantage of the two dimensional structure of panel data to construct a lattice specification. Then, a class of nonparametric, positive semidefinite covariance matrix estimators that allow for a general form of spatial dependence characterized by a metric of economic distance is introduced. This framework is applied to estimating spatial patterns in the residential demand for drinking water. Estimation results indicate that accounting for spatial dependence yields efficient estimate of the asymptotic variance matrix. Compared to non-spatial strategies, spatial dependence implies higher standard errors for all parameter estimates so as to strongly modify patterns of significance.

JEL Classification: C14; C23; D10

Keywords: Lattice models; Spatial dependence; Nonparametric covariance matrix estimation; Residential demand for water

*Tel.: (+33) 390 242 092; fax: (+33) 390 242 071; e-mail: `azomahou@cournot.u-strasbg.fr`

1 Introduction

The focus on spatial dependence has occurred in a range of fields in economics, not only in urban, real estate and regional economics, where the importance of location and spatial interaction is fundamental, but also in public economics, agricultural and environmental economics, and industrial organization. Recent studies concern among others the interplay between infrastrucure, investment and development (Seitz and Conrad, 1997, Seitz, 2000), the study of responses of real wages to local and aggregate unemployment rates over time (Ziliak et al., 1999) and the estimation of a hedonic model for residential sales transactions (Bell and Bockstael, 1999).

From a consumer analysis side, the presence of variables and errors which are fundamentally spatial in character is natural. The formation of preferences, through observation and replication of neighbors' behavior (habit formation), may lead to the presence of a spatially correlated dependent variable. At the same time, the availability of substitute goods, the dissemination of information, and soil and climate conditions may all be unobservable variables which are potentially spatially correlated and which contribute to spatial correlation in demand errors.

This paper presents and estimates empirically a lattice model using household panel data on water consumption.[1] Specifically, it is argued that a spatial model of dependence between observations can be useful to model dependence among economic agents. Each agent's observation is modelled as a realization of a random process at a point in an Euclidean space: a random field.[2] The distance between two agents in this space may reflect their proximity or similarity with respect to individuals' observables as well as unobservables, a notion of "economic distance" or "social distance".

Arguments for constructing dependence structure based on the economic distance between agents have been suggested in several studies. For instance, Hautsch and Klotz (1999) argued that geographic distance becomes less and less important while individuals and firms pay more and more attention to those being in the same comparable situation. Then, the occurrence of spatial correlation is not restricted to geographical spaces: observations can be thought of as being located in an abstract space, with some socioeconomic characteristics (such as per capita income or percentage of the population in a given racial or ethnic group etc.) being the dimensions. Those observations are then

[1] To illustrate lattices let us consider a spatial process $\{Z_\tau : \tau \in \mathcal{S}\}$ where $\mathcal{S} \subseteq \mathbb{R}^d$ is an index set of a countable collection of regularly or irregularly scattered spatial sites and these sites are supplemented with a neighborhood structure. Neighborhood structure is generally modeled either by a connectivity matrix (say W_n, where W_n is a $n \times n$ matrix, with elements $w_{ij} = 1$ if sites i and j are juxtaposed, $w_{ij} = 0$ if not; n is the number of sites) or by a graph-theoretic formalism (the sites become vertices, which are connected with edges for contiguous objects). Such spatial processes are called lattices. See, e.g., Cressie (1991) for a taxonomy of spatial data structures.

[2] See also e.g., Case (1991) and Driscoll and Kraay (1998) for applications of random fields.

said to be similar. This approach has been used successfully, among others, by Case, Rosen, and Hines (1993). Another example is the study of Conley and Ligon (1995) on growth regression, where the measurement of economic distance used to compute spatial standard errors is a measurement of transportation cost of physical capital between countries. It should be noted that the economic distance measurement is particularly useful in case data are collected at an individual level (households) or micro-level data for which geographic distance (in a physical sense) cannot be computed.

In this study, we combine the economic distance framework and the empirical complexity that follows from the use of generalized method of moments (GMM) in estimating lattice models. The joint distribution of random variables at a set of points is a function of the economic distances between them. This modelling strategy allows for a simple characterization of considerable interdependence among agents. It incorporates a more complex dependence across individuals than models with group-specific effects, e.g., Moulton (1990), or with scalar indexed dependence, e.g., Domowitz and White (1984). However, it requires that the econometrician has information regarding this economic distance.

As demonstrated theoretically in Conley (1999), GMM estimators remain consistent with such dependency but their asymptotic distribution theory and subsequently, covariance matrix estimation procedure and efficient GMM estimators are different from the time-series situation. The approach adopted here is to estimate variance matrices using nonparametric methods that allow for spatial dependence. The time-series analogue is that followed by, e.g., Newey and West (1987), Andrews (1991), White (1984), or Domowitz and White (1984). Covariance matrix estimation is presented in the situation when the economic distances between observations are measured exactly. In that case, a class of consistent positive semidefinite (p.s.d.) estimators of the asymptotic variance matrix is introduced.

The empirical purpose of the study consists in estimating spatial patterns in the residential demand for drinking water using households' micro-level data. The data has been collected bi-annually from 1994 to 1997 over one thousand of people. The presence of spatial dependence in the analysis of water demand is usually attached to availability of water resources as well as habit formation. In this context, the specification used may be viewed as a model of endogenously changing tastes, which permits to check for social interdependence by testing the extent to which households look to a reference group when making water consumption decisions. It may also be thought of as indicating the magnitude and the direction of interactions between consumers with respect to the availability of water resources. For instance, Priscoli (1999) reported that a river basin and watershed has among the most persistent examples of how the functional and spatial necessities of water can form consumers' preferences. Indeed, the spatial and functional characteristics of a river basin influenced human settlement and interaction long before the idea of the river basin started to be formalized into legal and administrative terms.

As mentioned above, the construction of a neighborhood structure is based on the notion of economic distance. Here, we use households' income. Although such an indicator should be considered carefully, for the sample concerned, this is the best indicator we have at hand to measure the similarity between households.[3] Moreover, when using data on individuals, we cannot define a physical distance in a geographical sense between them. As a result, the impact of allowing for spatial dependence will bear upon inference, not parameter estimates. The spatial problem will then be to correct for the standard errors of parameter coefficients for spatial correlation based on location. Thus, we will focus both parameter estimates and compare the standard errors for various variance estimators.

For the sample concerned, we find that accounting for spatial dependence implies higher standard errors for all parameter estimates so as to strongly modify patterns of significance. We also observe the following finding. One of the main features in the data is that of differentiated cold and hot water meters by household. Such a fine water consumption recording should prompt households to save water. Estimation results indicate that this is not so.

The remainder of the paper is organized as follows. Section 2 provides the theoretical background. We give an overview of the model and present the estimation strategy. The specification is a statistical model of spatial dependence where agents live on a lattice. The large sample results for GMM estimators and variance matrix estimators using spatially dependent data are stated. Section 3 presents the data. Since the data are new and have necessitated an important collection work, we give a detailed description of the salient features following from the sampling procedure as well as descriptive statistics. Section 4 presents the empirical implementation and discusses estimation results. The policy implications are also discussed. Concluding remarks are given in Section 5.

2 Theoretical framework

The specification is a modified version of the model developed by Conley (1999). We first state the data structure underlying the model. Then, we present the estimation procedure based on GMM estimators. Finally, we discuss the issue related to the estimation of the asymptotic variance matrix.

[3]As pointed out by Anselin and Bera (1998), the use of socioeconomic indicators to measure economic distance may pose problems for poorly chosen economic determinants.

2.1 Model and data generation process

The starting point of the analysis is a population of individuals (for example households) which is assumed to reside in a Euclidean space, with each individual i located at a point s_i. The investigator's sample consists of realizations of agents' random variables at a collection of locations $\{s_i\}$ inside a sample region. The latter is assumed to be a sequence of finite closed, convex regions $\{\Lambda_\tau\}$ which increases in area as $\tau \to \infty$. The population of potentially observable locations forms a lattice with irregular spacing denoted H. The sample consists then of $H \cap \Lambda_\tau$. In order to increase the sample size as $\tau \to \infty$, Λ_τ is assumed to grow uniformly in at least two non-opposing directions.[4] Attached to each position s is a vector X_s of random variables.

In a given sample, the econometrician's data consists of two parts. The first contains the realization of X_{s_i} at all points s_i within the set Λ_τ, i.e. $\{X_{s_i} : s_i \in \Lambda_\tau\}$. The number of points N_τ in Λ_τ is a random variable. The second part of the data is a $N_\tau \times N_\tau$ symmetric matrix D, with d_{ij} denoting an element of D, that is, the distance between points s_i and s_j.

The population at hand is assumed to reside at integer coordinate locations with known economic distances. Then, the positions s_i of individuals in the sample can be inferred from the interpoint distances up to a normalization of location and orientation. A method of finding coordinates of points in an Euclidean space given their interpoint distances is among others the well known Multidimensional Scaling (MDS).[5] Because the econometric theory here does not depend on these two normalizations, locations are then observed.

The data generating process has two components. The first part is a sampling process that determines which individuals and hence which locations s_i are observed. The values of attributes of individual i are determined by X_s. Following Clark (1973), the observable X_{s_i} on the lattice H, is said to be subordinated to X_s and the process generating H is called the directing process. We consider these concepts in turn.

Let $Z^2 = \{(i,t)|i = 1, \cdots, N;\ t = 1, \cdots, T\}$ be the two-dimensional lattice of integers with i and t denoting respectively the individual and the time subscript. The value of random variables at each location s_i is determined by a $\ell \times 1$ random field X_s with the index $s \in Z^2$. The X_s is assumed to be stationary, i.e. the joint distribution of X_s for any collection of indices s is invariant to a shift in the indices. X_s is also assumed to be mixing. The mixing condition will be defined later. The directing process determining H can be described by a regular lattice indexed random field W_s that is equal to one if location s is sampled and zero otherwise, with expectation $E(W_s) = \lambda$

[4]This assumption is made to ensure that indexing by a dependence vector is not superfluous.

[5]See e.g., Maital (1978) and DeSarbo, Kim, and Fong (1999) for examples of application and further details on the MDS.

and is assumed to be stationary and mixing.

The directing process W_s can be viewed as arising from both the distribution of the population on the lattice and the actual survey sampling scheme used in collecting the data. For example, if the population of agents were uniformly distributed on the plane and the sample was an independent sample of agents then W_s would be an i.i.d. dummy variable. If, however, the population was concentrated in certain parts of the plane and agents were sampled independently then W_s would exhibit positive spatial correlation. However, spatial correlation in W_s could also arise from cluster sampling of an evenly distributed population. Thus it is not possible to distinguish sampling schemes from population distributions given locations alone.

Conley (1999) considers the following mixing condition. Let $\mathcal{F}_\Lambda$ be the σ-algebra generated by a given random field ψ_{s_m} with $s_m \in \Lambda$, where Λ is a compact set, and $|\Lambda|$ is the number of $s_m \in \Lambda$. Let $\phi(\Lambda_1, \Lambda_2)$ denote the minimum Euclidean distance from an element of Λ_1 to an element of Λ_2. The mixing coefficient is defined as

$$\alpha_{k,l}(n) = \sup_{(F_1 \in \mathcal{F}_{\Lambda_1}, F_2 \in \mathcal{F}_{\Lambda_2})} \{|P(F_1 \cap F_2) - P(F_1)P(F_2)|\}, \tag{1}$$

$$|\Lambda_1| \leq k, |\Lambda_2| \leq l, \phi(\Lambda_1, \Lambda_2) \geq n.$$

The definition of mixing for X_s and W_s requires $\alpha_{k,l}(n)$ for each process to converge to zero as $n \to \infty$ at a rate which will be specified below.

2.2 Estimation strategy

The pattern of dependence is that where the distance between agents' positions, corresponding to their economic distances, characterizes the dependence between their random fields. If two agents' locations s_i and s_j are close, then their X_{s_i} and X_{s_j} may be very highly correlated. As the distance between s_i and s_j increases, the random variables X_{s_i} and X_{s_j} become closer to being independent in the sense made precise above. In the following, we provide conditions for the consistency and asymptotic normality of the GMM estimator when the X_{s_i} are dependent, as well as tractable p.s.d. covariance matrix estimators.

We assume that economic theory has produced a moment condition involving X_{s_i} which can be used to estimate a parameter vector of interest, θ_0 identified as the unique solution of the moment condition:

$$Eg(X_{s_i}, \theta_0) = 0, \tag{2}$$

where E is an expectation operator with respect to the true distribution of X_{s_i}; $g : \mathbb{R}^\ell \times \mathcal{B} \to \mathbb{R}^\nu$; θ_0 is a $k \times 1$ vector of the true unknown parameters to be estimated and θ is in the interior of $\mathcal{B}$, a compact subset of $\mathbb{R}^k$; ν is assumed greater than or equal to k, so there are $\nu - k$ over-identifying restrictions. A GMM estimator $\hat{\theta}_\tau$ of θ_0 is obtained as

$$\hat{\theta}_\tau = \arg\min_{\theta \in \mathcal{B}} \left[\frac{1}{N_\tau} \sum_{i=1}^{N_\tau} g(X_{s_i}, \theta) \right]' S_\tau \left[\frac{1}{N_\tau} \sum_{i=1}^{N_\tau} g(X_{s_i}, \theta) \right], \tag{3}$$

where $S_\tau \overset{\text{a.s.}}{\to} S_0$, a positive semidefinite weighting matrix. If S_τ is a consistent estimator of the inverse of the asymptotic covariance matrix of the moment conditions, we obtain efficient GMM estimates of θ_0, see, e.g., Hansen (1982). Let us consider now the asymptotic properties of $\hat{\theta}_\tau$.

A set of sufficient conditions for consistency is analogous to the set of conditions assumed by Hansen (1982) in the time-series case. The only difference between the current spatial model and the time-series case is that a pointwise law of large numbers for random fields rather than one for time series is used.

Assumption 1 *(A1:i) Λ_τ grows uniformly in two non-opposing directions as $\tau \to \infty$; (A1:ii) S_τ converges in probability towards S_0, a positive-definite matrix; (A1:iii) X_s and W_s are mixing; $g(.,\theta_0)$ is Borel measurable for all $\theta \in B$ and $g(X;.)$ is continuous on B for all $x \in \mathbb{R}^\ell$, and first moment continuous on B.*

Proposition 1 (Conley, 1999) *Given conditions (A1:i)-(A1:iii), $\hat{\theta}_\tau \to \theta_0$ in probability as $\tau \to \infty$.*

The asymptotic distribution of $\hat{\theta}_\tau$ follows from the mean value expansion of $g(X_{si}, \hat{\theta}_\tau)$ around θ_0. We assume that for each component of $g(.)$, there exists $\bar{\theta}_i$ such that

$$g_i(\hat{\theta}_\tau) = g_i(\theta_0) + \frac{\partial}{\partial \theta'_\tau} g(\bar{\theta}_i)(\hat{\theta}_\tau - \theta_0), \tag{4}$$

where $\bar{\theta}_i$ has elements between $\hat{\theta}_\tau$ and θ_0. The mean value expansion yields the expression:

$$\begin{aligned}\sqrt{N_\tau}(\hat{\theta}_\tau - \theta_0) = - &\left\{ \left[\frac{1}{N_\tau} \sum_{i=1}^{N_\tau} \frac{\partial}{\partial \hat{\theta}'_\tau} g(X_{s_i}, \hat{\theta}_\tau) \right]' S_\tau \left[\frac{1}{N_\tau} \sum_{i=1}^{N_\tau} \frac{\partial}{\partial \hat{\theta}'_\tau} g(X_{s_i}, \hat{\theta}_\tau) \right] \right\}^{-1} \\ &\times \left[\frac{1}{N_\tau} \sum_{i=1}^{N_\tau} \frac{\partial}{\partial \hat{\theta}'_\tau} g(X_{s_i}, \hat{\theta}_\tau) \right]' S_\tau \frac{1}{\sqrt{N_\tau}} \sum_{i=1}^{N_\tau} g(X_{s_i}, \theta_0).\end{aligned} \tag{5}$$

The conditions for consistent estimation of the first portion in the right-hand side of the expression (5) that converges in probability, are essentially the same as in the time-series case. Regularity conditions for the second portion of that has an asymptotic distribution, that is

$$\frac{1}{\sqrt{N_\tau}} \sum_{i=1}^{N_\tau} g(X_{s_i}, \theta_0), \tag{6}$$

are slightly different from the time-series case. A central limit theorem due to Bolthausen (1982) for stationary mixing random fields on regular lattices can be used. The expressions for limit theorems used in Conley (1999) are simplified by defining a process that is observed at all points within the sample region Λ_τ so that there are no missing observations. Let a new process $Y_s(\theta)$ take on the following value at a point s:

$$Y_s(\theta) = \begin{cases} g(X_s, \theta) & \text{if } W_s = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Observe that partial sums of $Y_s(\theta)$ equal partial sums of $g(X_s, \theta)$. Expressing $(1/\sqrt{N_\tau})$ $\sum_{i=1}^{N_\tau} g(X_{si}, \theta_0)$ in terms of $Y_s(\theta_0)$ yields:

$$\frac{1}{\sqrt{N_\tau}} \sum_{i=1}^{N_\tau} g(X_{s_i}, \theta_0) = \frac{\sqrt{|\Lambda_\tau|}}{\sqrt{N_\tau}} \frac{1}{\sqrt{|\Lambda_\tau|}} \sum_{s \in \Lambda_\tau} Y_s(\theta_0), \tag{8}$$

where the notation $|\Lambda_\tau|$ refers to the number of integer lattice locations falling within Λ_τ. A limit distribution for (6) is obtained by applying Bolthausen's central limit theorem to $(1/\sqrt{|\Lambda_\tau|}) \sum_{s \in \Lambda_\tau} Y_s(\theta_0)$ and showing that $\sqrt{|\Lambda_\tau|}/\sqrt{N_\tau}$ converges in probability (Bolthausen, 1982). Additional moment and mixing conditions required to obtain a limiting distribution for $\hat{\theta}_\tau$ are the following.

Assumption 2 *(A2:i)* $\sum_{m=1}^{\infty} m\alpha_{k,l}(m) < \infty$ *for* $k+l \leq 4$*; (A2:ii)* $\alpha_{1,\infty}(m) = o(m^{-2})$*; (A2:iii) for some* $\delta > 0$*,* $E(\|g(X_s, \theta_0\|)^{2+\delta}$ *and* $\sum_{m=1}^{\infty} m\alpha_{1,l}(m)^{\delta/(2+\delta)} < \infty$*; (A2:iv)* $(\partial/\partial\theta) g(X_{s_i}, \theta)$ *is Borel measurable for all* $\theta \in B$*, continuous on* B *for all* $X \in \mathbb{R}^\ell$*, and first moment continuous;* $E[(\partial/\partial\theta_0) g(X_{s_i}, \theta_0)]$ *exists and has full rank; (A2:v) defining* $V = \sum_{s \in Z^2} EY_0(\theta_0), Y_s(\theta_0)'$*,* V *is a non-singular matrix.*

X_s and W_s are assumed to satisfy the mixing and moment conditions $(A2:i)-(A2:iii)$. Conditions $(A2:iv)$ and $(A2:v)$ ensure that the expected derivatives are consistently estimated and expectations of derivatives and variance matrices have full rank. These conditions imply the following claim.

Proposition 2 (Conley, 1999) *If conditions* $(A1:i)-(A1:iii)$ *and* $(A2:i)-(A2:v)$ *are satisfied, then*

$$\sqrt{N_\tau}(\hat{\theta}_\tau - \theta_0) \xrightarrow{d} N\left(0, D_0' \lambda^{-1} D_0\right), \quad \text{as } \tau \to \infty$$

where $\lambda = E(W_s)$ *and*

$$D_0' = \left\{ E\left[\frac{\partial}{\partial\theta_0'} g(X_{s_i}, \theta_0)\right]' S_0 E\left[\frac{\partial}{\partial\theta_0'} g(X_{s_i}, \theta_0)\right] \right\}^{-1} E\left[\frac{\partial}{\partial\theta_0'} g(X_{s_i}, \theta_0)\right]' S_0. \tag{9}$$

2.3 Covariance matrix estimation

The asymptotic covariance matrix is based on the moment conditions and is estimated as

$$\Omega = \lambda^{-1} V = \lambda \sum_{s \in Z^2} E Y_0(\theta_0) Y_s(\theta_0)'. \tag{10}$$

To describe how the estimation procedure works, it is convenient to refer to the two coordinates in s directly. So let $s = [m_1, m_2]$. Also take the region Λ_τ to be a rectangle so that $m_1 \in \{1, 2, \cdots, M_1\}$ and $m_2 \in \{1, 2, \cdots, M_2\}$ where the dependence of M_1 and M_2 on τ is suppressed to ease notation. The sum in (10) can be interpreted as a sum of spatial autocovariances, analogous to the time-series case where the infinite sum of an autocovariance function is the asymptotic variance matrix of sample averages. A spectral representation of covariance stationary processes on the plane is also available, allowing the interpretation of the covariance matrix in (10) as a spectral density at frequency zero.

Familiar discrete time-series spectral density estimation techniques generalize to random fields on integer lattices.[6] Smoothed periodograms can also be used to estimate two-dimensional spectra; see, e.g., Priestley (1989) for details on this topic. For the special cases where $Y_{m_1,m_2}(\theta_0)$ is Gaussian or is a linear process, smoothed periodogram spectral density estimators have been proven to be consistent by Rosenblatt (1978). However, the above restrictions on X_s, g and W_s do not imply that $Y_{m_1,m_2}(\theta_0)$ is Gaussian or linear. Furthermore, the fact that the parameter θ_0 is unknown and has to be replaced by the estimate $\hat{\theta}_\tau$ must be addressed.

Let us define L_{M_1} and L_{M_2} as lag truncation points for M_1 and M_2 respectively. The class of variance matrix estimators we consider is that formed by taking weighted averages of spatial autocovariance terms with weights that are zero for points farther than L_{M_1} and L_{M_2} apart in each direction. This class is analogous to time-series spectral density estimators whose time domain weights equal zero after a cut-off lag. The Bartlett window estimator used by Newey and West (1987) and the truncated estimator in White (1984) are two such time-series covariance matrix estimators.

Let us consider the following estimator $\hat{V}$ of V, constructed as a weighted average of products of Y_{m_1,m_2} terms:

$$\begin{aligned}\hat{V} =& \frac{1}{M_1 M_2} \sum_{j=0}^{L_{M_1}} \sum_{k=0}^{L_{M_2}} \sum_{m_1=j+1}^{M_1} \sum_{m_2=k+1}^{M_2} K_{M_1 M_2}(j,k) [Y_{m_1,m_2}(\theta_0) Y_{m_1-j,m_2-k}(\theta_0)' \\ &+ Y_{m_1-j,m_2-k}(\theta_0) Y_{m_1,m_2}(\theta_0)'] \\ &- \frac{1}{M_1 M_2} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} Y_{m_1,m_2}(\theta_0) Y_{m_1,m_2}(\theta_0)',\end{aligned} \tag{11}$$

[6] See, e.g., Yaglom (1987) for an extensive discussion of spectral representations of random fields.

for an array of weights $K_{M_1 M_2}(j,k)$ that are uniformly bounded and are such that $K_{M_1 M_2}(0,0) = 1$ and $K_{M_1 N_2}(j,k) \to 1$ as $M_1 \to \infty$ and $M_2 \to \infty$. The subtraction of the second term is simply to remedy double counting of the $j = 0$ and $k = 0$ term in the first sum. With known economic distances such an array is simple to construct, a truncated weighting array with weights equal to one for $j < L_{M_1}$, $k < L_{M_2}$ is one example.[7] As detailed below, a careful choice of weighting arrays will produce covariance matrix estimates that are necessarily p.s.d.

If θ_0 were known, then we could use $(N_\tau/|\Lambda_\tau|)^{-1} = (N_\tau/(M_1 M_2))^{-1}$ to estimate λ^{-1} and estimate Ω by $(N_\tau/(M_1 M_2))^{-1}\hat{V}$. This estimator is of course not feasible because $Y_{m_1,m_2}(\theta_0)$ is not observed as θ_0 is unknown. The simple solution is to plug in the estimator $\hat{\theta}_\tau$, for θ_0 yielding the feasible estimator of Ω:

$$\begin{aligned}\hat{\Omega}_\tau =& \frac{1}{N_\tau} \sum_{j=0}^{L_{M_1}} \sum_{k=0}^{L_{M_2}} \sum_{m_1=j+1}^{M_1} \sum_{m_2=k+1}^{M_2} K_{M_1 M_2}(j,k)[Y_{m_1,m_2}(\hat{\theta}_\tau) Y_{m_1-j,m_2-k}(\hat{\theta}_\tau)' \\ &+ Y_{m_1-j,m_2-k}(\hat{\theta}_\tau) Y_{m_1,m_2}(\hat{\theta}_\tau)'] \\ &- \frac{1}{N_\tau} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} Y_{m_1,m_2}(\hat{\theta}_\tau) Y_{m_1,m_2}(\hat{\theta}_\tau)'.\end{aligned} \tag{12}$$

A set of sufficient conditions for consistent estimation of Ω are:

Assumption 3 *(A3:i) The $K_{M_1 M_2}(j,k)$ are uniformly bounded, and $K_{M_1 M_2}(j,k) \to 1$ as $\tau \to \infty$ that is $(M_1, M_2) \to \infty$; $L_{M_2} = o(M_2^{-1/3})$ and $L_{M_1} = o(M_1^{-1/3})$; (A3:ii) for some $\delta > 0$, $E(\|g(X_s, \theta_0\|)^{4+\delta} < \infty$ and for both the X_s and W_s processes the mixing coefficient $\alpha_{\infty,\infty}(m_1)^{\delta/(2+\delta)} = o(m_1^{-4})$; (A3:iii) $E \sup_B \|Y_{m_1,m_2}(\theta)\|^2 < \infty$ and $E \sup_B \|(\partial/\partial\theta)[Y_{m_1,m_2}(\theta)]\|^2 < \infty$.*

These conditions include restrictions on the weighting array in condition $(A3:i)$; a strengthening of moment and mixing assumptions in $(A3:ii)$, and in $(A3:iii)$, a set of conditions on g and its derivatives to ensure consistency despite using $\hat{\theta}_\tau$ in place of θ_0.

Proposition 3 (Conley, 1999) *Given conditions $(A1:i)-(A1:iii)$, $(A2:i)-(A2:v)$, and $(A3:i)-(A3:iii)$, $\hat{\Omega}_\tau$ converges to Ω in probability as $\tau \to \infty$.*

[7] As pointed out by Conley (1999), precise distance information is not needed to obtain a consistent covariance matrix estimate. If distance information were only known up to broad categories then a weighting array $K_{M_1 M_2}(j,k)$ that was constant over the distances within a category would be feasible with this imprecise distance information. This estimator will not always be p.s.d., unfortunately, since the spectral window corresponding to the step function space domain window (its Fourier transform) will be negative in some regions.

The weights $K_{M_1M_2}(j,k)$ can be chosen so as to guarantee p.s.d. point estimates. One way to construct such weights involves using a spectral representation of the stationary process $Y_{m_1,m_2}(\theta_0)$ that is analogous to those for time series. V is proportional to the spectral density matrix at frequency zero of the process $Y_{m_1,m_2}(\theta_0)$ and the estimator $\hat{V}$ can be expressed as a weighted periodogram spectral density estimator. The periodogram is p.s.d., so if the weights $K_{M_1M_2}(j,k)$ correspond to a non-negative spectral window then the estimate will be p.s.d. Weights satisfying these conditions can be easily made by the product of usual time domain weights.

Consider for example the two-dimensional weight function that is a Bartlett window in each dimension:

$$K_{M_1M_2}(j,k) = \begin{cases} \left(1 - \frac{|j|}{L_{M_1}}\right)\left(1 - \frac{|k|}{L_{M_2}}\right) & \text{for } \ |j| < L_{M_1},\ |k| < L_{M_2}, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Its Fourier transform is non-negative, so with these weights $\hat{\Omega}_\tau$ will be p.s.d.

3 Data and variables

In this section, we describe and illustrate in detail the main features of the data. Others considerations are given in the Appendix.

The data we use in this study are the only ones collected from the French network of residential drinking water distribution. They are provided by the "*Office Public des Habitations à Loyers Modérés (OPHLM)*" (the public agency for council flats) of the municipality of Sarreguemines.[8] Data were recorded from 1994 to 1997. Except for the first semester of 1995 when they are quarterly, all other observations have been collected half-yearly. The sample represents an unbalanced panel of about 1000 households.

In general, two types of documents were used to collect the information we needed. The first document summarizes indications related to water meters, indices of water consumption and the actual dwelling area measured in m2. The number of the water meter identifies a given flat. At this stage, the main information is provided by the consumption indices. Indeed, consumption values, expressed in cubic meter (m3) are derived directly from the difference between two consecutive indices. When a household leaves a residence, the meter is switched back to zero before a new entrance.

The second type of document we exploited is termed "lodging identification sheets". In this document, households are asked to describe their situation, e.g. their income, their employment situation, the demographic structure of the family, their marital status, etc. Unfortunately, we do not have any information on their appliances. Some

[8]The municipality of Sarreguemines is located in the Moselle department of in the North-East of France.

of these indicators (typically household characteristics) are used to construct dummy regressors. We now consider the main features of the sampling scheme.

3.1 Water consumption

The OPHLM of Sarreguemines manages several blocks of flats for which it also provides drinking water. These residences may be one, two or more-room flats or separate houses. Since 1994, the OPHLM of Sarreguemines has adopted a new method for continuous recording of households' consumption of water. This method is based on the "remote reading technology" that makes it possible for example to separate the cold water consumption from the hot water consumption. Residences built since 1994 have been then systematically provided with individual water meters, whereas those built before 1994 are gradually fitted out. The total population of households concerned by these individual meters is about 70% of the total residences managed by the agency. Water consumption from other residences is also available but only at an aggregate level. Those residences are not included in our sample. Among households having these meters at their disposal, about 40% are provided with two meters at least. Each of these meters records information (real water consumption, water leaks, etc.) for each type (cold or hot) of water consumption.

Thus, some households have differentiated water meters. Others have only one water flow meter gathering at the same time the hot water consumption and the cold water consumption. That is explained in the following way. On the one hand, the households whose hot water supply is ensured by a collective production of hot water have at least two meters. In this case, the water invoice is also differentiated and the price elements (the price of m3 and the share of the subscription) of each type of water appear clearly on the invoice. For these households, one has the differentiated values for cold water and hot water consumption. On the other hand, the households whose hot water supply is ensured by an individual heating device do not have differentiated meters. These households have one water flow meter covering at the same time the hot water and the cold water consumption. For example, they are households with individual boilers who thus heat their own water. In this case, we do not have the differentiated values of consumption. In such a situation, the water bill does not mention the hot water elements. For several reasons which will become clear later, the current study is concerned only with the total consumption of water, i.e., the sum of cold water and hot water. Also, we will define a dummy variable in order to check the impact of differentiated water meters on demand.

The technique of remote reading facilitates the exploitation of these meters and makes it possible to receive sufficiently precise information to study the demand for water. The objective of the OPHLM is to save water by decreasing households' consumption. The method implemented to achieve this is, on the one hand, the installation of individual meters which allow the individualization of water bills and, on the other hand, remote monitoring which makes it possible to detect leaks and over-consumption, and thus to detect possible problems.

For each block of flats with households equipped with individual meters, there is also an aggregate measurement of water consumption. This aggregate measure is carried out with another unique meter which is different from the individuals meters. We will explain the relevance of this aggregate measurement later. For the moment, it should be noted that the water consumption readings from the overall meters (aggregate measurement) are carried out by the "Compagnie Générale des Eaux" (CGE), whereas the readings of consumption on the individual meters are carried out by the "Société Loralsace de Contrôle et de Gestion (SLCG)". As we will see below, this double counting, made on behalf of the OPHLM of Sarreguemines, makes it possible to calculate an indicator called "network connection coefficient" that relates aggregate readings and the sum of individual readings.

The standard of living of the selected population varies slightly. Among them, some households have a very low disposable income and are often in arrears with their water bills. We do not know the proportion of households in this situation. However, Table 1 reports the history of the delayed payment of bills. We notice that the maximum arrears does not exceed one month. See also the Appendix for details concerning problems raised by the price of water.

Insert Table 1 here

The OPHLM of Sarreguemines has two documents which summarize the water consumption of the households: water invoices of the CGE and the readings of the SLCG. The water invoices contain aggregate information on water consumption, total water expenditure and pricing. Meter reading established by the SLCG carries individual information about consumption and the characteristics of the apartments. Previously, the tenants paid for water in proportion to their dwelling area. Now they pay for their consumption of water. This can encourage them to limit their consumption and to have repaired water leakage quickly. Indeed, it was noticed that the immediate consequence of the installation of individual water meters was an important decrease of the water consumption in the two years which followed.

The system of remote reading provides automatically, every hour, the statement of the meters. It is enough in a given period to find one hour during which consumption is nil to be certain that there is no leakage. Conversely, a water flow meter recording continuously is an indicator of water leakage. The meters can detect very low flows: 10 liters per hour with a precision of more or less 5%. This is enough to measure the flow of a thin filament of water (16 liters per hour) or of a flushing system which leaks (25 liters per hour).

The meter reading that the SLCG transmits to the OPHLM summarizes in a large table the consumption of water in cubic meter for a six month period. The organization of the data is hierarchical going from the largest (group of buildings) to the smallest (apartments). In a first part of the table, each tenant is identified by a number referring to the apartment which he or she occupies. The following information is given: the building, the number of the staircase and the number of the floor. The tenant of an apartment is identified from the date of arrival and departure. He or she also provides information about the flat (F2 for two rooms, F3 for three rooms, etc., are thus coded). Information regarding meters are also noted. The equipment in water meters varies from one building to another. It is sometimes restricted to only one cold water meter, but often comprises four meters: a hot water meter and a cold water meter in the kitchen, and the same equipment in the bathroom. The meters are identified by a number. The second part of the table establishes the indices of consumption. For each meter, we observe biannual indices, except the year 1995 for which we have quarterly indices. Water volumes are obtained by substractiion between two consecutive indices.

For some residences, the OPHLM initiated collective water heating in spite of the increase in hot water consumption which resulted. There are about 80 residences equipped with individual boilers. Previously, the tenants, anxious to reduce the amount of the hot water invoices that they paid to "Gaz de France", did not heat their home sufficiently (and in some cases not at all). This caused damp patches and the appearance of mold in the dwellings. Consequently, the OPHLM adopted a collective heating system, which resulted in an increase of hot water consumption. Indeed, from this moment, the hot water bills are included in the charges of the dwelling and are no longer perceived on the basis of invoices as previously established by "Gaz de France".[9]

So far we have outlined that there also exists an aggregate measurement of water consumption from a different meter to the individual ones. It is important to note that for a given block of flats, the sum of individual consumptions never equals the value provided by the aggregate consumption meter. The aggregate consumption always exceeds the sum of individual ones. This is due to extra-consumption: for example the water used by equipment common to a building. From these two measurements of water consumption, the Agency computes a coefficient termed "network connection coefficient" which serves as an overvaluation factor of household consumption. This

[9] It is not a question of the perception of the invoices in the subjective sense of the agents, but of the perception of an amount of money. However, one may think that the new mode of payment of the hot water invoices induced some subjectivity by consumers, who are not any more directly in touch with their hot water expenses. This expenditure remains nevertheless identifiable for the consumers who wish to obtain this information.

indicator ensures the link between the collective invoices of CGE and the individual ones of the SLCG. For example, in 1997, the "network connection coefficient" for a given building was 1.053. Then the water consumption of all households living in this building will be overvalued by 5.3% in the water bills. This is not always easily accepted by consumers.

3.2 Income

The data concerning households' income come from a census carried out by the OPHLM from Sarreguemines for the years 1994, 1995 and 1996. Each household was invited to provide two documents: a "lodging identification sheet" and a copy of their tax record. They include the following elements:

- An exhaustive census of the tenants asking for each of them: name, date of birth, sex, mention of a possible handicap and rate of disability which would result from it, family tie with the holder of the lease, nationality, profession, employer, and finally monthly income and family benefits.
- Further information concerns the holder of the lease: his social identification number and marital status.
- Details of other sources of income such as maintenance, housing allowance, personalized help with housing, services coming from the "Caisse d'Allocations Familiales" such as the "Revenu Minimum d'Insersion", family allowances and "allocation parent isolé").

The tax record and the lodging identification sheet do not refer to the same year. We will see the resulting problems. The information recovered is the tenant's annual income plus that of the spouse and, if relevant, that of the children. It should be noted that not all tax cuts correspond to a decrease of income. Some are to be regarded as choices of consumption, others as arbitrations between consumption and saving.

The first category covers renovation and repairs carried out in the residence of the taxpayer or gifts to "charities". In the second category all the contingency contracts (complementary insurances, old-age insurances) are to be found as well as saving plans. Paid alimonies on the other hand correspond to income reductions and are considered as such. In general, the income data raises several problems. The principal reservation relates obviously to the reliability of the statement of income. The tax records are established on the basis of income tax returns. All the tenants do not provide their tax record. Some lose it, others do not wish that one knows in a precise way their income and state that they are tax free.

Several difficulties were encountered when studying the identification sheets. Indeed, very often, the information they comprise does not fit the taxation record. One of the reasons is the two-year delay between the identification sheet and the tax record. The population of the OPHLM is in general young and mobile and moves frequently. The tenants sometimes provide income tax returns corresponding to a time when they did not yet occupy their apartment. Moreover, the situation of the same individual can vary rapidly. From one census to another, one of the parents can lose his (her) employment, be entitled or not to social benefits, and as happens also sometimes, find another job. For this same individual, the wages are thus prone to strong variations from one census to another. So, in order to obtain an income variable for our empirical analysis, we gather all income information available together, i.e. the sum of the non-salaried income and the salaried income.

3.3 Households' characteristics

Households' characteristics are mainly computed from the lodging identification sheets. As noticed previously, in this document households are asked to describe their situation including their income and source, their employment situation, the demographic structure of the family, their marital status, their date birth etc. We use this information to define some characteristics such as: age variables, profession of each person in the family, total number of persons and number of children in the household, marital status of parents, nationality etc.

On the one hand, for each household we have the family connections between the members of the family, the profession of each person, their nationality, etc. when the information is mentioned. On the other hand, the data on the phenomenon under study, i.e. the demand for water, is provided at the household level. So, we have to aggregate some of these characteristics to create usable variables. For example, the basic data appears with the profession of each person in the family. For this very micro-level information, we have enumerated about a hundred of professions. In this case, we have used the official nomenclature of socio-professional classification to define seven categories of profession. See INSEE (1994) for the official nomenclature. We have also aggregated the nationality information to define four nationality classes. The individual date of birth and the family connections are combined to count the total number of children in a family as well as the number of children under eighteen and over eighteen. See Appendix (Table 10) for a definition of all variables.

Insert Table 2 here

Insert Table 4 here

Insert Table 5 here

A summary of descriptive statistics is given in Tables 2, 4 and 5. National statistics indicate an average water consumption tendency around 120 cubic meters per household and per year. These figures vary from one household to another. When we compare these indicators with those computed from our sample, we notice that the average consumption is of the same magnitude. Nevertheless minima are surprising. Indeed, for the total water consumption, except for 1994.2 and 1996.1, all periods are associated with a minimum of 1 cubic meter. At each period, about ten households are concerned with these low values that do not result from measurement errors. This may be due to the fact that only integer values of consumption are accounted for by the OPHLM. But we do not have any explanation for these low values other than perhaps a long period of absence of a single family. We also observe from Table 5 that French people, married people and categories of profession CSP-5 to CSP-8 are the most represented in the sample. Again see Appendix (Table 10) for a complete definition of variables included the categories of professions.

4 Estimation results and discussion

In this section, we first describe the empirical procedures as well as estimation figures. Then, we discuss these results.

4.1 Empirical implementation

The data has two main characteristics which make it possible to apply the theoretical framework sketched in Section 2. We have an unbalanced panel of households supplemented by a group structure. The group structure is represented by blocks of flats; about seventy are considered. In terms of the model described in the theoretical framework, Λ_τ can be viewed as the geographic area of residences. The hypothesis that Λ_τ grows to increase the sample size as $\tau \to \infty$ corresponds both to gradually equipping buildings constructed before 1994 with individual water meters as well as to constructing new blocks of flats from 1994, which then, are automatically equipped with individual water meters. The two alternatives allow us to increase the sample size. Moreover, the sampling process is such that we know in which block of flats each household is located, and what floor they are on.

The directing process can be considered as corresponding to the distribution of the population which is concentrated in certain parts of a plane: several households in a given block of flats. Moreover, there is a cluster sampling in that for each block of flats, we are interested only in households equipped with individual water meters. All these reasons motivate the use of a spatial framework. With this information at hand, we first organize the data so as to form a lattice; then we compute the economic distance between households as follows.

Since panel data is two dimensional by nature (one dimension pertaining to the individual and the other to time), we take advantage of this and of the group structure of the sample (block of flats) to reorganize the data as follows. Firstly, take the time dimension $t = 1, \cdots, T$. To each wave, associate the same number of individuals in a given group. Then repeat this calculation until obtaining a remainder number of individuals, say, $\bar{N}$ less than T. Remember that here, $T = 7$. To complete the organization, withdraw $\bar{N}$ from the sample. Indeed, unless you are lucky, all the individuals which are present in the original panel will not be retrieved in the reorganized sample. Consequently, a few observations ($\bar{N} < T$) will be eliminated. This is not a serious drawback however because the individual dimension of a panel is usually larger than the time dimension. It is known that for $N > T$, the asymptotic is usually based on N. We then think that the elimination of a few observations will not modify this asymptotic. It should be noticed from the reorganized sample that the two dimensions of the panel are preserved. At the end, we have a lattice, the two dimensions of which can be viewed as the associated coordinates.

The next step was to compute the economic distance between households. The measurement we have used here is households income. The construction of the metric is based on the multidimensional scaling algorithm of Mardia et al. (1979, pp. 394-423). See the Appendix for a brief description of this algorithm. See also e.g., Maital (1978) for alternative algorithms. To give an intuition of the procedure, let us say that the general notion of distance underlying the multidimensional scaling may not correspond directly to a Euclidean coordinate system. This procedure allows us to take any set of distances between points and use the multidimensional scaling algorithm to obtain a set of coordinates in a Euclidean space whose interpoint distances approximate the original distance between points. Once we have a set of point coordinates we can proceed with estimations.

Two types of estimation are conducted.[10] First we estimate a panel data model using fixed effects and random effects specifications. Then, we carry out estimations on the lattice model.

First, we use the group structure of the data to test for individual correlation in standard error estimates without modelling spatial dependence explicitly. Such an approach can give an idea about the presence of spatial patterns. Table 6 shows results of

[10]STATA and GAUSS procedures to implement the calculations of this paper are available from the author upon request. Due to restrictions on the dissemination of the data, their use requires permissions from the OPHLM of Sarreguemines.

an estimation from pooled OLS with robust estimator of variance.[11] The first part of the table presents pooled OLS estimates without accounting for group structures. The second part shows pooled estimates with group clusters in standard errors. That is, we specify that observations are independent across groups but not necessarily independent within groups. Observe the difference in the two robust standard error estimates. It seems that there is a cluster effect which may also be indicative of the existence of spatial effects. Some variables remain significant on the two estimates, such as the dummy of having a differentiated water meter for the cold and the hot water consumption, the number of people in a family. For the professional dummies, only CSP-7 (i.e., retired; see the Appendix (Table 10) to remember the definition of professions) remains significant whereas for the nationality dummies, only one (African) becomes insignificant. Note however that the number of significant variables decreases strongly, fourteen in the non-cluster case, and only eight in the cluster case.

Insert Table 6 here

Observe that there is no price variable in the estimation. Again, see the Appendix for issues concerning the price of water. There are several reasons for this. At first, too many ambiguities characterize its determination. We are unable to obtain from the OPHLM of Sarreguemines a current water price. Secondly, from the information in Table 3, and assuming that the fixed parts are really considered by households as a component of the price of water, one may be tempted to compute an average price since the marginal price will be the same for all households. But, in this case, the only variation in the average price will come from variations in consumption. Finally, the OPHLM does not consider the fixed part as a component of the price of water as we have pointed out in the description of data. Consequently, we do not include any price information in the estimation.

Insert Table 7 here

In Table 7, we present fixed effects (OLS on the within regression) and random effects estimates using GLS. That is we consider estimating an error component model of the form $y_{it} = \alpha + x_{it}\beta + \nu_i + \varepsilon_{it}$, where ν_i is an individual specific effect; it differs between units but, for any particular individual, its value is constant. ε_{it} is an idiosynchratic error term with zero mean, and uncorrelated with x_{it} and ν_i and homoskedastic. Whatever the properties of ν_i and ε_{it}, if the relation is assumed to be true, it must also be true that $\bar{y}_i = \alpha + \bar{x}_i\beta + \nu_i + \bar{\varepsilon}_i$, where $\bar{y}_i = \sum_t y_{it}/T$, $\bar{x}_i = \sum_t x_{it}/T$ and $\bar{\varepsilon}_i = \sum_t \varepsilon_{it}/T$. Substracting $y_{it} - \bar{y}_i$ provides the basis for the within regression, that is the fixed-effects

[11]Here, the term robust is taken in the sense of Huber/White/sandwich estimator of variance.

estimator. Variables for which no estimates are available in Table 7 are those which are time invariant because of their elimination due to the above difference. The random effect estimator follows from the hypothesis of random ν_i and uses the assumption of no-correlation.[12]

Insert Table 8 here

Insert Table 9 here

In Table 8 we present OLS estimates and standard errors calculated in two different ways. First, the heteroskedasticity consistent standard errors are in the column labeled White std.err. (HET). Then, standard errors computed using the estimator $\hat{\Omega}_\tau$ defined in (12) that allows for spatial dependence as well as heteroskedasticity are in the column labelled SP (for spatial). In computing these spatial standard errors, we use the two-dimensional Bartlett window defined in equation (13). We may expect that the same qualitative results obtained here can be achieved with other kernels such as the truncated or the spectral.[13]

The OLS estimator can be considered as a just-identified GMM estimator. However, the impact of allowing for dependence will likely be even greater in overidentified systems where estimated asymptotic variance matrices will determine weighting matrices and hence parameter estimates and tests for overidentifying restrictions as well as standard errors. This is done in GMM estimation allowing for spatial correlation. The results are reported in Table 9. Again, we used the two-dimensional Bartlett window defined in equation (13). For comparison, we have also reported estimates from two stages least squares assuming spatial independence.

To check the validity of relation (2), that is the null H_0 that the moment condition underlying the GMM is verified, we used the GMM criterion function test of overidentifying conditions; see, e.g., Gouriéroux and Monfort (1991, pp. 619-631) for details. From the estimated $\hat{\Omega}_\tau$ and the assumptions of the model, the test statistic is

$$\xi_n := \left[\frac{1}{N_\tau}\sum_{i=1}^{N_\tau} g(X_{s_i}, \hat{\theta}_\tau)\right]' \hat{\Omega}_\tau \left[\frac{1}{N_\tau}\sum_{i=1}^{N_\tau} g(X_{s_i}, \hat{\theta}_\tau)\right],$$

where $\hat{\theta}_\tau$ is the vector of GMM estimates. Under the null, we have $\{\xi_n \geq \chi^2_{(1-\alpha)}(\nu-k)\}$. The statistic is computed to be 6.898. As a result, given the associated degree of freedom, there is no rejection of the null. Here again the standard errors for the spatial GMM estimator are found to be higher than those from the non-spatial 2SLS.

[12]One advantage of the fixed effects model is that, since we are conditioning on ν_i, we do not need to assume that they are independent of the regressors. However, the random effects model will yield more efficient estimates when it is appropriate. Mundlak (1978) and Chamberlain (1984) are classical references on issues concerning the relationship between the fixed effects and the random effects models.

[13]See Andrews (1991) for several widely used windows.

4.2 Discussion

The estimates in Table 8 and 9 provide evidence that allowing for spatial dependence can be important for conducting inference. Overall, the magnitude of the difference between spatial standard errors and HET standard errors are enough to change the values of the t-tests for significance and to impact the confidence interval enough to change the economic significance of the estimated parameters.

Table 8 and 9 illustrate another important point. Usually it is assumed that spatial dependence does not imply that standard errors will rise. Such results have been found to be empirically valid. For example, Conley (1999) examined empirically whether growth rates are related to measurements of human capital, political stability, fiscal variables etc., using as economic distance the cost of transporting physical capital between countries. Based on the same spatial methodology, he found that most of the standard error estimates corrected for spatial dependence are smaller than their non-spatial counterparts. He argued that the asymptotic variances may be smaller with spatially dependent data, just as asymptotic variances can be lower for dependent time-series. Here we found something opposite. The correction of standard errors for parameter coefficients for possible spatial dependence based on location leads systematically to higher standard errors.

In general, what can we learn through this study about household demand for drinking water? First of all there is evidence of spatial dependence. As we have noticed earlier, this result supports the idea that regionalized behavior in households' consumption of drinking water matters in some French municipalities. Such a behavior may also be linked to the availability of water resources. The spatial pattern here may also be thought of as indicating the effect of no intra-individual variation in water prices. Variations in this variable may depend on the distance between households.

We notice that the sign of the coefficient of the income variable is always negative. In general this coefficient is significant for estimates where spatial patterns are not taken into account, whereas it is not significant for spatial estimates. Maybe this is due to the omission of a price indicator. Some characteristics also appear as determinant factors of the demand. For example the average age of a household, the number of persons by household and the dwelling area are significant and the associated coefficients carry the expected sign. The significance of profession and nationality dummies varies according to estimations.

Previously, we emphasized that some households in our sample are equipped with two water meters or more, whereas others have only one meter. In order to study the effect of meter characteristics on the demand, we defined a dummy for having at least two meters. The result is surprising. Indeed, whatever the estimation results are, the dummy variable of water meter is highly significant and positive, which means that the differentiation of water meters results in an increase of the demand. Remember that the target of the OPHLM of Sarreguemines in equipping households in differentiated individual water meters was to have them save water. Relying on our estimation results, we can say that this objective is not necessarily attained. This result can be interpreted tentatively as follows.

One may think that having a fine recording of water consumption may reduce it. This seems not to be the case: one explanation is that the values of each consumption pattern (hot and cold consumption) that the system of remote reading will record, will often be lower than the two combined. As a result, households will keep in mind the differentiated figures and will react to them rather than react to the aggregate figure. This may lead to a wrong perception of their actual water consumption. Thus, it would be of interest to equip households with only one remote reading system which would provide them with the two components of their consumption as well as an aggregate figure. It would be also of interest to clarify the pricing of water and to include it in the remote reading system.

5 Conclusion

This study presents and estimates empirically a lattice model using households level panel data on demand for water. The model allows us to make the most of information on households' interdependence to characterize spatial dependence structures. We use households' income to get such a measurement of economic distance between them, though a measurement of their "similarities". Then, a MDS algorithm is used to retrieve the coordinates of each household in a plane. These coordinates are used in the estimation procedure to correct for spatial dependence in estimated standard errors based on locations.

We showed empirically that accounting for such dependence impacts strongly the standard error estimates for all parameters so as to modify levels of significance. As a result, the spatial dependence allows us to obtain consistent standard error estimates. We have also noticed that the remote reading system that equips households does not necessarily urge them to reduce their consumption.

A fundamental restriction placed on the structure of dependence is that it can be characterized by a configuration of points in an Euclidean space, presumably lower than the sample size. Moreover, estimation is carried out under the assumption of exact economic distance. That is there is no measurements error on the characteristics of households which are used to compute the neighborhood relation. There are many potential measurements of the economic distance between agents that are imperfect. When economic distances are measured with error, additional information may be needed on the distribution of measurement errors. Even if this complicates the estimation procedure, it seems to be a promising direction for future empirical studies.

Acknowledgements

Part of this study was conducted when I was visiting the Center for Operations Research and Econometrics (CORE), Université Catholique de Louvain-La-Neuve, Belgium. I am heartily grateful to François Laisney for his detailed constructive criticisms and continued encouragement in developing this work. I am also indebted to Helmut Seitz for helpful comments and discussions. I would like to thank especially Timoty Conley for clarifying some aspects of his study. I thank the "*Office Publique des Habitations à Loyers Modérés (OPHLM) de Sarreguemines*" for providing me with the data used in this study. Financial support of the "*Ecole Nationale du Génie de l'Eau et de l'Environnement de Strasbourg (ENGEES)*", (National School of Water and Environment Engineering of Strasbourg), is gratefully acknowledged. All errors and shortcomings remain my own.

Appendix A: Issues concerning the price of water

Water pricing is described in two documents: the tariffing of CGE (overall water sale) and the tariffing of the OPHLM of Sarreguemines.

The invoices sent by CGE to the OPHLM of Sarreguemines are not individual invoices but relate to buildings, even sets of buildings. They are calculated on real and estimated water consumption. A new invoice model was adopted in 1994. It describes in a clearer way than before the elements which compose the price of water. The first part relates to the supply of water. It distinguishes between the "syndicate" tax and what concerns the supplier (a two part tariff: a fixed part which does not depend on the quantity and a marginal price). A second part reflects the costs of collection or water treatment. The third part gathers the taxes and royalties levied by the National Fund for Water Conveyance (FNDAE), for the water agencies and the control of pollution. Among these elements, one can see the total expenditure which incorporates all the elements that compose the pricing and that is concerned with the CGE, i.e. the price of the cubic meter and the fixed part.

The main difficulty encountered in the use of tariff information is that the network of meters changed over time. It is thus sometimes difficult to know to which building a meter refers. The aggregations are not the same from one year to another, which makes the comparisons difficult. Curiously, the addresses indicated on the invoices are not always reliable. For example, a building supposed to be located in "rue des Rossignols" is in fact located in "rue des Fauvettes". Or when the information about this building is missing from "rue des Rossignols" and another appears in "rue des Hirondelles", they may be the same building. We might therefore consider the numbering of the meters, which remains constant. In addition, individual data consists in indications of consumption, whereas the price of water appears only in the invoices, which represent collective consumption. As a result, an increase in the price of water will result only in an increase of the collective charges. The consumer will not be aware that he is saving when repairing, for example, a tap which leaks.

The water invoices of the OPHLM are all calculated on real consumption. This tariffing scheme has changed during the years. We distinguish a variable part (the price of a cubic meter or the marginal price) according to whether it concerns cold water or hot water and a fixed part corresponding to the cost of the reading, the network maintenance etc. The fixed part remains identical for the counting of cold water and hot water. Curiously, the fixed part is not considered by the OPHLM as a component of the price of water, even if it is clearly identified as such in the charges, the details of which is not provided to the tenants. Table 3 gives some relative information on the price of water we have succeeded in identifying.

Insert Table 3 here

In general, due to the lack of precision and many ambiguities, we are not able to derive a price of water other than a marginal price, i.e., the price of the cubic meter as computed by the OPHLM of Sarreguemines. However, this price is the same for all the households in a given period and varies very little during a this period of time.

Appendix B: Variables used in the estimation

Insert Table 10 here

Appendix C: Multidimensional scaling algorithm

Multidimensional scaling (MDS) is concerned with the problem of constructing a configuration of n points in Euclidean space using information about the distances between the n objects. The interpoint distances themselves may be subject to error. The distances need not be based on Euclidean distances, and can represent many types of dissimilarities between objects. Also in some cases, one may start not with dissimilarities but with a set of similarities between objects.

Definition 1 *A reasonable measure of similarity, $s(A, B)$, should have the following properties:*

i) $s(A, B) = s(B, A)$,
ii) $s(A, B) > 0$,
iii) $s(A, B)$ *increases as the similarity between* A *and* B *increases.*

Definition 2 *A distance matrix* D *is called Euclidean if there exists a configuration of points in some Euclidean space whose interpoint distances are given by* D*; that is, if for some* p*, there exists points* $(x_1, \cdots, x_n) \in \mathbb{R}^p$ *such that*

$$d_{rs}^2 = (x_r - x_s)'(x_r - x_s).$$

Theorem 14.2.1 in Mardia (1979, p. 397) enables us to tell whether D is Euclidean, and if so, how to find a corresponding configuration of points.

Suppose we have given a distance matrix D which we hope can approximately represent the interpoint distances of a configuration in a Euclidean space of low dimension k; usually $k = 1, 2, ...$ The matrix D may or may not be Euclidean; however, even if D is Euclidean, the dimension of the space in which it can be represented will usually be too large to be of practical interest. Choose the configuration in $\mathbb{R}^k$, the coordinates of which are determined by the first k eigenvectors of B, where

$$B = HAH, \qquad A = (a_{rs}), \quad a_{rs} = -\frac{1}{2}d_{rs}^2,$$

and $H = I - n^{-1}ee'$, $e = (1, \cdots, 1)$. If the first k eigenvalues of B are "large" and positive and the other eigenvalues are near 0 (positive or negative), then hopefully, the interpoint distances of this configuration will closely approximate D. This configuration is called the classical solution to the MDS problem in k dimensions. It is a metric solution.

For computational purposes we shall summarize the calculations involved:

1. From D construct the matrix $A = (-1/2)d_{rs}^2$.

2. Obtain the matrix B with elements $b_{rs} = a_{rs} - \bar{a}_{r.} - \bar{a}_{.s} + \bar{a}_{..}$ where

$$\bar{a}_{r.} = \frac{1}{n}\sum_{s=1}^{n} \bar{a}_{rs}, \quad \bar{a}_{.s} = \frac{1}{n}\sum_{r=1}^{n} \bar{a}_{rs}, \quad \bar{a}_{..} = \frac{1}{n^2}\sum_{r,s=1}^{n} \bar{a}_{rs}.$$

3. Find the k largest eigenvalues $\lambda_1 >, \cdots, > \lambda_k$ of B (k chosen ahead of time), with corresponding eigenvectors $X = (x_{(1)}, \cdots, x_{(k)})$ which are normalized by $x_{(i)}'x_{(i)} = \lambda_i$, $i = 1, \cdots, k$. (We are supposing here that the first k eigenvalues are all positive.)

4. The required coordinates of the points P_r are $x_r = (x_{r1}, \cdots, x_{rp})'$, $r = 1, \cdots, k$, the rows of X.

References

ANDREWS, D. W. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, (59–3), 817–858.

ANSELIN, L., AND A. BERA (1998): "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," *in Handbook of Applied Economic Statistics, A. Truchmuche (eds.)*, pp. 237–289.

BELL, K. P., AND N. E. BOCKSTAEL (2000): "Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data," *The Review of Economics and Statistics*, (82–1), 72–82.

BOLTHAUSEN, E. (1982): "On the Central Limit Theorem for Stationary Mixing Random Fields," *The Annals of Probability*, (10), 1047–1050.

CASE, A. (1991): "Spatial Patterns in Household Demand," *Econometrica*, (59), 953 –965.

CASE, A., H. ROSEN, AND J. HINES (1993): "Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States," *Journal of Public Economics*, (52), 285–307.

CHAMBERLAIN, G. (1984): "Panel Data," *in Handbook of Econometrics, A. Truchmuche and B. Machinchose (eds.)*, (2), 1248–1313.

CLARK, P. (1973): "A Subordinate Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, (41), 135–155.

CONLEY, T. (1999): "Generalized Method of Moments Estimation with Cross Sectional Dependence," *Journal of Econometrics*, (92), 1–45.

CONLEY, T., AND E. LIGON (1995): "Economic Distance, Spillovers and Growth," *Working paper University of California Berckley.*

CRESSIE, N. (1991): *Statistics for Spatial Data.* New York: Wiley-Interscience.

DESARBO, W., Y. KIM, AND D. FONG (1999): "A Bayesian Multidimensional Scaling Procedure for the Spatial Analysis of Revealed Choice Data," *Journal of Econometrics*, (89), 79–108.

DOMOWITZ, I., AND H. WHITE (1984): "Nonlinear Regression with Dependent Observations," *Econometrica*, (52–1), 143–161.

Driscoll, J. C., and A. C. Kraay (1998): "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," *The Review of Economics and Statistics*, (80–4), 549–560.

Gouriéroux, C., and A. Monfort (1989): *Statistiques et Modèles Econométriques, Vol.1–2*. Paris: Economica.

Hansen, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, (50), 1029–1054.

Hautsch, N., and S. Klotz (1999): "Estimating the Neighborhood Influence on Decision Makers: Theory and an Application on the Analysis of Innovation Decisions," *University of Konstanz*.

INSEE (1994): *Nomenclatures et Codes, "Nomenclature des Professions et Catégories Socioprofessionnelles"*. Institut National de la Statistique et des Etudes Economiques.

Maital, S. (1978): "Multidimensional Scaling: Some Econometric Applications," *Journal of Econometrics*, (8–1), 33–46.

Mardia, K., J. Kent, and J. Bibby (1979): *Multivariate Analysis*. London: Academic Press.

Moulton, B. (1990): "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *The Review of Economics and Statistics*, (72), 334–338.

Mundlak, Y. (1978): "On the Pooling of Times Series and Cross-Section Data," *Econometrica*, (46), 69–85.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, (55–3), 703–708.

Priestley, M. (1989): *Spectral Analysis and Time Series*. Academic Press.

Priscoli, J. (1999): "Water and Civilization: Using History to Reframe Water Policy Debates and to Build a New Ecological Realism," *Water Policy*, (1), 623–636.

Rosenblatt, M. (1978): *Dependence and Asympotic Independence for Random Processes*. Washington: D. C. Mathematical Association of America.

Seitz, H. (2000): "Infrastructure, Industrial Development and Employment in Cities," *fothcoming in International Regional Science Review*.

Seitz, H., and K. Conrad (1997): "Infrastructure Provision and International Market Share Rivalry," *Regional Science and Urban Economics*, 27, 715–734.

White, H. (1984): *Asymptotic Theory for Econometricians.* London: Academic Press.

Yaglom, A. (1987): *Correlation Theory for Stationary and Related Random Functions, vols.I and II.* Springer, New York.

Ziliak, J. P., B. A. Wilson, and J. A. Stones (1999): "Spatial Dynamics and Heterogeneity in The Cyclicality of Real Wages," *The Review of Economics and Statistics*, (81–2), 227–236.

List of Tables to be inserted

Table 1: History of delayed payment of water bills

Features	Period			
	1994	1995	1996	1997
Proportion of unpaid bills	5.69%	6.65%	6.86%	5.78%
Number of days	21	24	25	21

Table 2: Descriptive statistics of water consumption*

Period	Total water consumption in m3 mean	std.	min.	max.	obs.	Hot water consumption in m3** mean	std.	min.	max.	obs.
1994.2	70.038	49.230	2	527	913	23.886	18.170	1	126	460
1995.1	33.415	22.910	1	166	809	12.775	10.338	1	85	495
1995.2	29.322	21.331	1	277	800	10.299	8.591	1	80	471
1995.3	61.603	43.923	1	413	870	20.198	16.902	1	165	503
1996.1	66.159	43.988	4	384	689	22.861	18.633	1	200	504
1996.2	61.636	40.228	1	349	704	19.926	15.943	1	143	503
1997.1	64.143	39.796	1	318	695	21.517	15.588	1	133	520
1997.2	61.156	39.343	1	361	698	18.798	14.702	1	133	570
	Cold water consumption in m3** mean	std.	min.	max.	obs.	**Dummy of water meter***** mean	std.	min.	max.	obs.
1994.2	51.917	32.691	2	243	460	0.504	0.500	0	1	913
1995.1	22.492	15.331	1	111	495	0.516	0.499	0	1	960
1995.2	21.477	17.052	1	259	471	0.495	0.500	0	1	953
1995.3	48.598	33.126	1	252	503	0.492	0.500	0	1	1024
1996.1	46.777	31.246	2	265	504	0.494	0.500	0	1	1022
1996.2	44.457	28.056	1	206	503	0.476	0.499	0	1	1061
1997.1	46.275	28.463	3	233	520	0.503	0.500	0	1	1040
1997.2	46.935	30.234	1	228	570	0.265	0.441	0	1	2161

* The number of observations varies due to the unbalanced nature of the panel.

** The statistics are computed for households for which the two records exists.

*** 1=at least two water meters, 0 otherwise.

Table 3: Some indications on elements of the price of water

Component of the price	Period 1994	1995	1996	1997
Variable part (cold water) in FF/m3	11.48	12.58	13.04	13.2
Variable part (hot water) in FF/m3	37.4	37.03	39.8	40.43
Fixed part FF/semester	148	148	147.5	101.5

Source: Values constructed from information of the OPHLM of Sarreguemines.

Table 4: Descriptive statistics

	1994				1995				1996			
Variable	mean	std.	min.	max.	mean	std.	min.	max.	mean	std.	min.	max.
Total income (FF/1000)	6.525	4.993	0.047	27.470	6.432	4.821	0.013	27.234	6.606	4.956	0.071	32.679
Dwelling area (m2)	70.910	17.680	30.000	139.000	70.800	17.560	30.000	139.000	69.990	18.360	27.000	139.000
# Persons/house	2.806	1.686	1	11	2.903	1.754	1	11	3.018	1.762	1	10
# Children ≤ 18	2.283	1.272	1	7	2.444	1.289	1	7	2.513	1.265	1	7
# Children > 18	1.715	0.966	1	5	1.580	0.884	1	5	1.617	0.843	1	4
Average age*	27.921	19.583	1	90	28.845	19.648	1	91	29.620	19.748	1	92

* Average age denotes the average age of a household.

Table 5: Table of frequency count of characteristics

	1994		1995		1996	
Variable	freq.	percent.	freq.	percent.	freq.	percent.
NATIONALITY						
Africa	135	7.88	151	6.44	104	6.40
Others	52	3.03	52	2.22	36	2.22
Europe	23	1.34	33	1.41	18	1.11
France	1504	87.75	2109	89.94	1466	90.27
Total	1714	100.00	2345	100.00	1624	100.00
MARITAL STATUS						
Married	294	48.36	343	43.20	236	43.87
Single	79	12.99	121	15.24	73	13.57
Cohabitation	40	6.58	80	10.08	51	9.48
Widowed	97	15.95	116	14.61	76	14.13
Divorced	73	12.01	99	12.47	76	14.13
Separated	25	4.11	35	4.41	26	4.83
Total	608	100.00	794	100.00	538	100.00
PROFESSION*						
CSP-2	38	4.84	63	6.60	42	6.52
CSP-3	46	5.86	44	4.61	27	4.19
CSP-4	51	6.50	51	5.35	26	4.04
CSP-5	103	13.12	146	15.30	105	16.30
CSP-6	204	25.99	259	27.15	177	27.48
CSP-7	205	26.11	212	22.22	163	25.31
CSP-8	138	17.58	179	18.76	104	16.15
Total	785	100.00	954	100.00	644	100.00

* CSP-2: Craftsmen, trademen and company head; CSP-3: managers and high intellectual professions; CSP-4: intermediate professions; CSP-5: employees; CSP-6: workers; CSP-7: retired; CSP-8: not gainfully employed.

Table 6: Pooled OLS estimates (Dependent variable: residential demand for water measured in cubic meters)

	Pooled OLS (without cluster)			Pooled OLS (with cluster)		
Variable	coef.	std.err.	t-stat.	coef.*	std.err.	t-stat.
Intercept	-31.29	5.03	-6.21	—	11.24	-2.78
Total income	-0.42	0.16	-2.54	—	0.31	-1.37
Dummy of water meter	8.99	1.16	7.75	—	2.37	3.78
Average age	0.02	0.06	0.35	—	0.11	0.19
Number of Persons	14.92	0.60	24.64	—	1.23	12.08
Marital status	0.48	0.51	0.95	—	1.22	0.10
Dummy of sex	-4.89	1.74	-2.80	—	3.88	-1.26
Dwelling area (m2)	0.39	0.05	7.27	—	0.14	2.82
Professions (dummies)						
CSP-2	-6.65	3.21	-2.07	—	5.97	-1.11
CSP-3	-7.33	4.12	-1.77	—	6.64	-1.10
CSP-4	-2.21	3.72	-0.59	—	7.54	-0.29
CSP-5	-5.97	2.35	-2.54	—	4.38	-1.36
CSP-6	-4.92	2.42	-2.03	—	4.25	-1.16
CSP-7	-9.13	2.46	-3.71	—	4.09	-2.23
CSP-8	-3.58	2.69	-1.33	—	4.55	-0.78
Nationality (dummies)						
Africa	10.43	3.97	2.62	—	8.07	1.28
Europe	41.52	6.11	6.79	—	13.23	3.14
France	28.94	2.77	10.42	—	5.11	5.66
Others	29.29	6.23	4.70	—	14.06	2.08
Number of clusters			70			
Number of observations			2609			

* (—) Same coefficient estimates.

Table 7: Fixed and random effects estimates (Dependent variable: residential demand for water measured in cubic meters)

	Fixed effects* (Within estimator)			Random effects (GLS estimator)		
Variable	coef.	std.err.	t-stat.	coef.	std.err.	t-stat.
Intercept	—	—	—	-0.28	16.910	-0.02
Total income	-0.18	0.23	-0.78	-0.64	0.19	-3.33
Dummy of water meter	10.88	2.72	4.01	9.08	2.10	4.32
Average age	-3.33	0.35	-9.57	-0.33	0.11	-2.91
Number of persons	9.71	1.13	8.59	12.75	0.82	15.60
Marital status	—	—	—	1.02	0.90	1.13
Dummy of sex	—	—	—	-2.91	3.62	-0.80
Dwelling area (m2)	—	—	—	0.55	0.10	5.44
Professions (dummies)						
CSP-2	—	—	—	-5.65	4.74	-1.19
CSP-3	—	—	—	-3.06	5.68	-0.54
CSP-4	—	—	—	-4.48	7.47	-0.60
CSP-5	—	—	—	-5.10	3.29	-1.55
CSP-6	—	—	—	-5.31	2.64	-2.01
CSP-7	—	—	—	-4.47	3.13	-1.43
CSP-8	—	—	—	-2.38	2.38	-1.01
Nationality (dummies)						
Africa	—	—	—	-5.37	14.55	-0.37
Europe	—	—	—	17.35	15.43	1.12
France	—	—	—	8.20	13.72	0.59
Others	—	—	—	3.96	18.87	0.21

* (—) The within transformation wipes out time invariant parameters.

Table 8: OLS estimates of the lattice model (Dependent variable: residential demand for water measured in cubic meters)

Variable	coef.	std. HET*	t-stat HET	std. SP**	t-stat SP
Intercept	-11.29	4.77	-2.36	8.68	1.30
Total income	-0.43	0.17	-2.52	0.24	-1.79
Dummy of water meter	8.31	1.40	5.93	1.86	4.49
Average age	-0.16	0.05	-3.20	0.07	-2.28
Number of persons	13.21	0.51	25.90	0.81	16.51
Marital status	0.88	0.45	2.00	0.78	1.12
Dummy of sex	0.05	1.53	0.03	2.66	0.01
Dwelling area (m2)	0.65	0.04	16.00	0.10	6.40
Professions (dummies)					
CSP-2	-8.71	3.33	-2.62	4.77	-1.82
CSP-3	-9.37	4.52	-2.07	5.74	-1.63
CSP-4	-5.29	4.83	-1.09	5.66	-0.93
CSP-5	-6.23	2.65	-2.35	3.45	-1.80
CSP-6	-6.63	2.45	-2.70	3.51	-1.89
CSP-7	-3.98	2.56	-1.55	3.22	-1.23
CSP-8	-2.47	2.47	-1.00	3.88	-0.63
Nationality (dummies)					
Africa	-13.52	3.41	-3.96	4.97	-2.72
Europe	17.47	4.43	3.94	10.28	1.69
France	5.02	2.26	2.22	3.22	1.55
Others	4.55	6.81	0.66	9.12	0.49
Number of observations			3276		

*HET means White heteroskedastic robust standard error estimates.

**SP means standard error estimates corrected for spatial correlation.

Table 9: 2SLS and GMM estimates of the lattice model (Dependent variable: residential demand for water measured in cubic meters)

	Spatial 2LS			Spatial GMM		
Variable	coef.	std.err.	t-stat.	coef.	std.err.	t-stat.
Intercept	-13.09	4.75	-2.75	-18.69	9.84	-1.90
Total income	-0.46	0.17	-2.59	-0.42	0.25	-1.65
Dummy of water meter	7.64	1.39	5.49	7.80	2.04	3.82
Average age	-0.18	0.05	-3.43	-0.19	0.08	-2.25
Number of persons	13.13	0.51	25.74	13.36	0.87	15.25
Marital status	0.65	0.44	1.47	1.23	0.83	1.48
Dummy of sex	-0.57	1.52	-0.37	0.16	2.74	0.05
Dwelling area (m2)	0.65	0.04	14.75	0.67	0.11	5.87
Professions (dummies)						
CSP-2	-8.36	3.32	-2.51	-8.42	5.60	-1.50
CSP-3	-8.13	4.51	-1.80	-6.66	8.59	-0.78
CSP-4	-4.93	4.83	-1.02	-4.69	6.08	-0.77
CSP-5	-5.85	2.64	-2.21	-5.24	3.61	-1.44
CSP-6	-6.00	2.45	-2.44	-6.69	3.63	-1.84
CSP-7	-4.48	2.56	-1.75	-3.89	3.60	-1.08
CSP-8	-2.26	2.47	-0.91	-3.38	3.86	-0.87
Nationality (dummies)						
Africa	-12.62	3.41	-3.70	-10.82	5.29	-2.04
Europe	17.21	4.43	3.88	13.60	10.62	1.28
France	5.79	2.25	2.57	6.36	3.45	1.84
Others	5.74	6.81	0.84	11.96	23.76	0.50
$\chi^2_{5\%}$(d.o.f. = 4)			6.89			
Number of observations			3276			

Table 10: List of variables

Variable	Definition
Water consumption	Bi-annual consumption of drinking water by household, in m3
Total income	Sum of the wage income and non wage income, in 1000 FF
Dwelling area	Actual dwelling area, in m2
# Persons/house	Number of persons by household
Average age	Average age of the family
Dummy of water meter	1=at least two water meters, 0 otherwise
DUMMIES OF PROFESSION	
CSP-1	Farmers and owners: they are not represented in the sample
CSP-2	Craftsmen, trademen and company head
CSP-3	Managers and high intellectual professions
CSP-4	Intermediate professions
CSP-5	Employees
CSP-6	Workers
CSP-7	Retired
CSP-8	Not gainfully employed
DUMMIES OF NATIONALITY	
Africa	1=African; 0 otherwise
Europe	1=European; 0 otherwise
France	1=French; 0 otherwise
Others	1=others nationalities, 0=otherwise
FACTOR OF MARITAL STATUS	
Marital status	1=married; 2=single; 3=cohabitation; 4=widowed; 5=divorced; 6=separated